SECURITY ANALYSIS OF SECURED2

ALLIED ASSOCIATES INTERNATIONAL

TECHNICAL REPORT

Submitted By:
Allied Associates International, Inc.
6801 Kennedy Road
Suite 302
Warrenton, Virginia 20187
www.a2ius.com

TABLE OF CONTENTS

1. INTRODUCTION

Allied Associates International (A^2I) has been tasked to investigate a collection of binary files representing cipher text output from an unspecified system using unspecified input. With these files, A^2I was to analyze and ascertain any characteristics revealing the origin, nature, composition or leakage of information that would aid in revealing the source system or content.

Standard encryption techniques rely on algorithms that take potentially regular input data and generate output without any measurable regularity or pattern. Furthermore, with some algorithms, the same input can result in completely different outputs based on arbitrary (but known) processing parameter such as sizing or cryptographic salt.

In addition to the encrypted payload portion, output file dimensions and formats provide clues for decryption and reconstruction of source data. For the purposes of decrypting data at delivery endpoints, file formats may provide information required for reconstruction, e.g., sequence number for concatenation, cryptographic parameters for salt values, block size, stream vs block processing, initialization vectors or even indications of the very algorithms used for encryption and hash functions.

The analysis techniques used in this study were designed to detect any regularity or repetition within and between files accordingly, to see if a header format might be present, counters might be incrementing or encryption was sufficiently random. Statistical analysis functions were performed to test data distribution as a measure of observable randomness. Signal processing and correlation techniques were employed to see if a seemingly random patterns were sequentially recurring anywhere in the data. And basic analysis focused on the arrangement, number and size of the files in order to glean any clues regarding encryption algorithms possibly in use.

2. BASIC ANALYSIS

2.1. DIRECTORY / FILE CHARACTERISTICS

Presented to the task were four subfolders of cipher text data with numeric designations:

Directory	No. Files	Bytes	Name range
3000	13	1,370,728	data.0.bin.new - data.12.bin.new
3009	23	2,536,192	data.0.bin.new - data.22.bin.new
3010	23	2,573,920	data.0.bin.new - data.22.bin.new
3011	44	4,867,744	data.0.bin.new - data.43.bin.new
Total	103	11,348,584	

The files within each subfolder are named consecutively. The naming of these files and directories in such a consecutive manner reveal a possible ordering for concatenation or certain requirements for stateful decryption from file to file or at least the order they are to be processed. It may also belie an ordering in plain text naming of originating data if alphabetically aligned.

The files total 11,348,584 bytes of memory. Based on file sizes, it appears the data can be construed as 8 byte/64 bit integer units for a total of 1,418,573 (64 bit) integer units.

The individual file sizes are intriguing. About half the files have a length of exactly 14,209 64-bit integers. Ignoring the last file of each of the 4 directories, the minimum file size is 12,763 64 bit integers. It is suspected that each of the 4 directories may correspond to an original source file. Data from that source file is read and encrypted into the files within the corresponding directory. Although there is some question regarding the different sizes in the cipher text files, it is not believed to be from compression due to extremely small variance with many of the files rendering the exact same length.

It is suspected that there are record boundaries within the file, and that each file represents a fixed number of records. The records could be slightly different sizes, and if a block encryption algorithm is used (e.g., DES in CBC mode), the records will be processed in blocks of 64 bits. Partial records are padded to form a full 64 bits for encryption.

One reason for doing record-level encryption: individual records can be decrypted and processed without having to decrypt the entire file. If there anything that would indicate the processing of encrypted records, one would expect a block encryption with an independent initialization vector (IV) per block, or an integer counter/record id being used as an IV, rather than a chained block ciphering (CBC) mode.

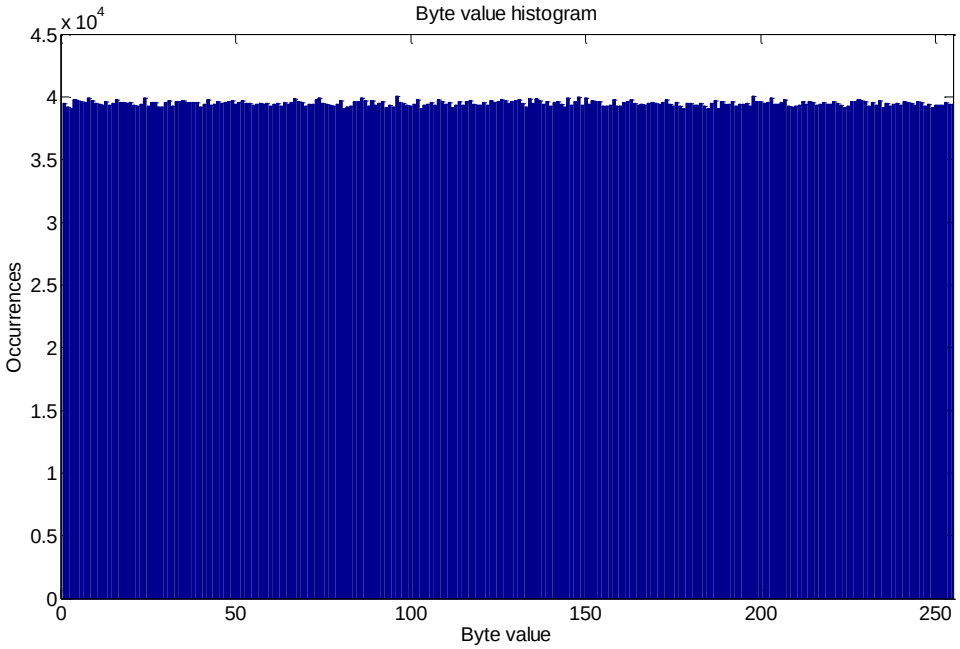
A simple program was used to compare each 64 bit integer units to all other 64 bit integers across all files, comparisons numbering on the order of 10^{12}. The comparison operation was a bitwise logical exclusive OR-ing of the values (XOR). All instances of pairs which differ in less than 8 bits were recorded. There were 30 such pairs among all comparisons considered. Examining these 30, there was nothing to distinguish them as in, for instance, a case of the low order bits incrementing. It was concluded that all of the data is encrypted or at least randomly masked.

2.2. STATISTICAL / SIGNAL ANALYSIS

The following set of tests and graphics were conducted using MATLAB, a scientific workbench from MathWorks.

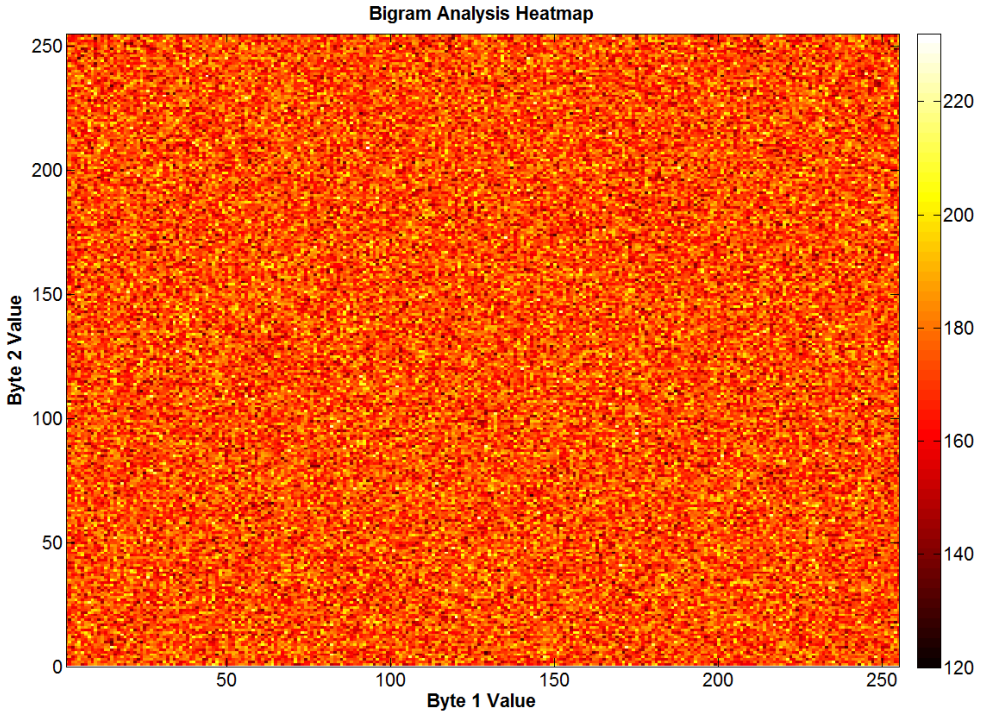
2.2.1. HISTOGRAM

This test measures the number of occurrences of each possible byte value in all files. Any significant deviation from a uniform distribution can severely impact the cryptographic strength of cipher text and suggests a weak encryption algorithm. In the sample data received, byte values are very uniformly distributed. All possible byte values are equally represented in the data and no significant deviations from the expected distribution can be observed.



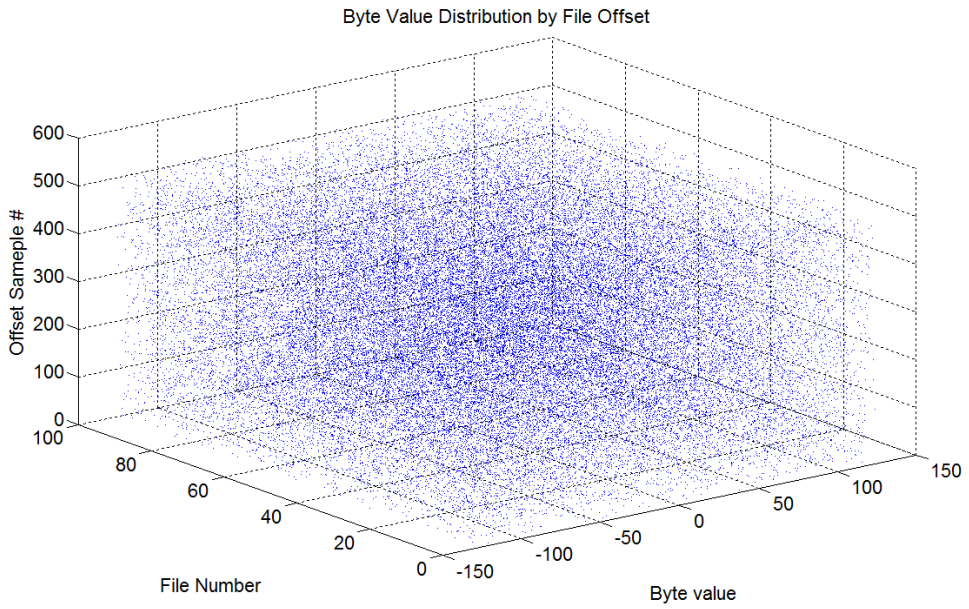
2.2.2. BYTE N-GRAM

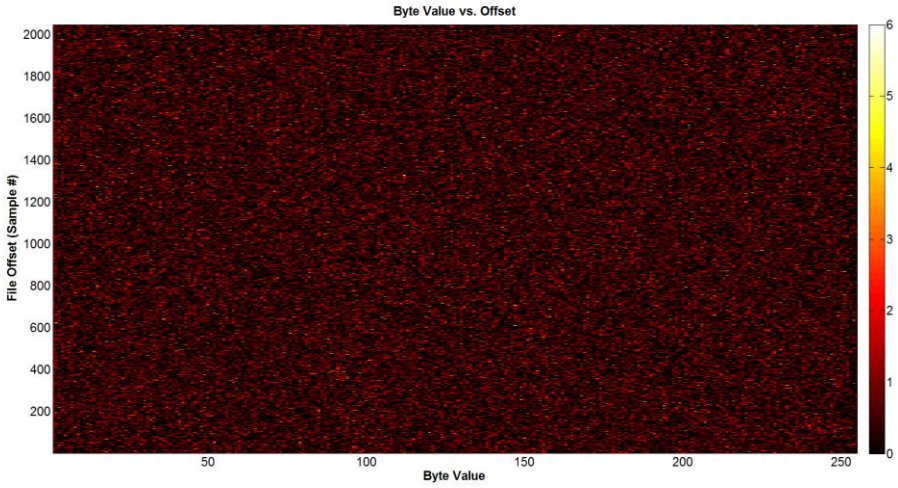
This test is meant to detect recurring two byte sequences in the data. While the histogram analysis of the data reveals that data is uniformly distributed and thus all byte values are equally represented it says nothing about byte sequencing. This test was conducted to detect any n = 2 byte motifs in the data that might be present if language were present or if encoding were limited to a subset of byte characters (e.g., base64 encoding). No significant patterns were discovered nor does anything appear to rise above the noise floor of the data.



2.2.3. BYTE VALUE VS. FILE OFFSET

This test measures byte value vs. file offset for each of the files individually. The test was designed to detect if there are byte values that appear more frequently in consistent regions (at various offsets from start) of the file. As evidenced by the following scatter plot, no such structures were found. Byte values are uniformly distributed throughout each file. In order to make the plot more visually informative, each file was randomly sampled to produce 512 sample elements, where the sample number correlates with an offset in the file. The test was also carried out on the entire, un-sampled data set with similar results.

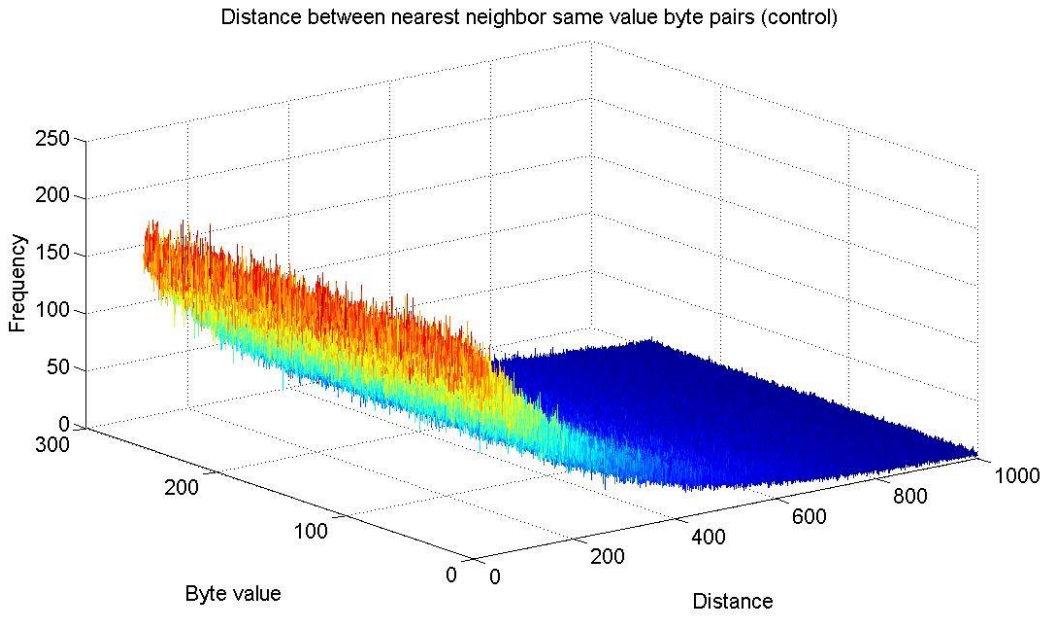




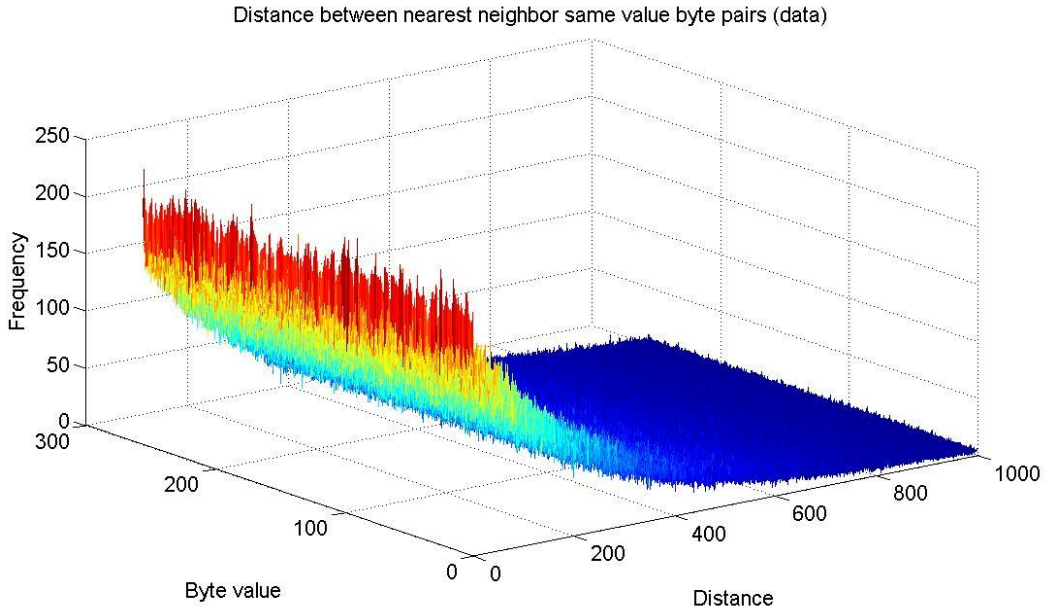
2.2.4. NEAREST SAME-VALUE BYTE PAIRS DISTANCES

This test measures the distance between same-valued byte pairs. It is designed to detect any periodically repeating byte values, such as any flags, or counters in data blocks. As a means of comparison, a large uniformly distributed random sample was generated and processed accordingly to produce the following surface.

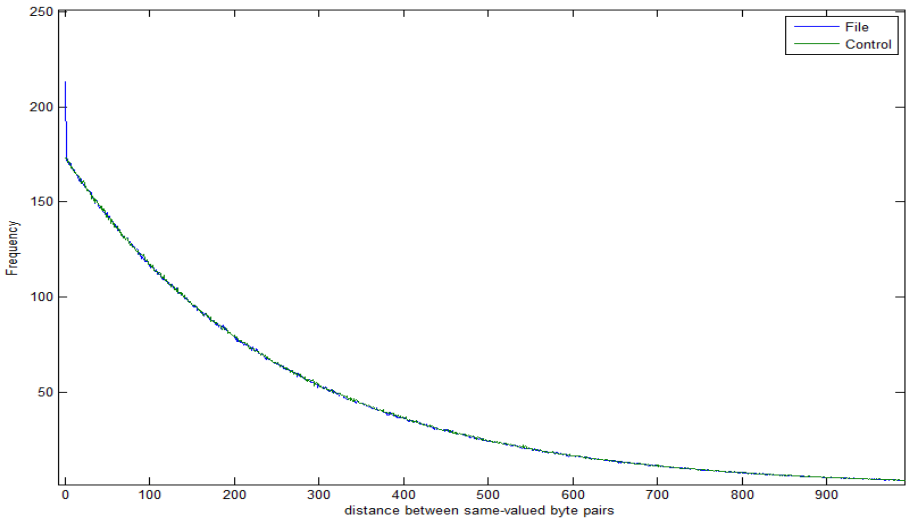
Randomly Generated Uniformly Distributed (control). This surface was generated using randomly generated data sample in MATLAB in order to compare to visualization using data from the cipher text.



Cipher text file data processed similarly produces this surface:



The test revealed that the distance between different same-value byte pairs in the data closely fits the expected distribution, with the exception that there are a higher number of adjacent same-value byte pairs (red ridge) when compared to the control sequence. The frequency of byte pairs with a distance of 1 is elevated—meaning that we have a higher number of consecutive same value bytes than one would expect from a truly uniform distribution. This holds true for all byte value [0-255] “doublets”. This is more readily visible in the following plot, which shows the mean number of occurrences of each distance across all byte values, essentially creating a two-dimensional projection of the surface above.

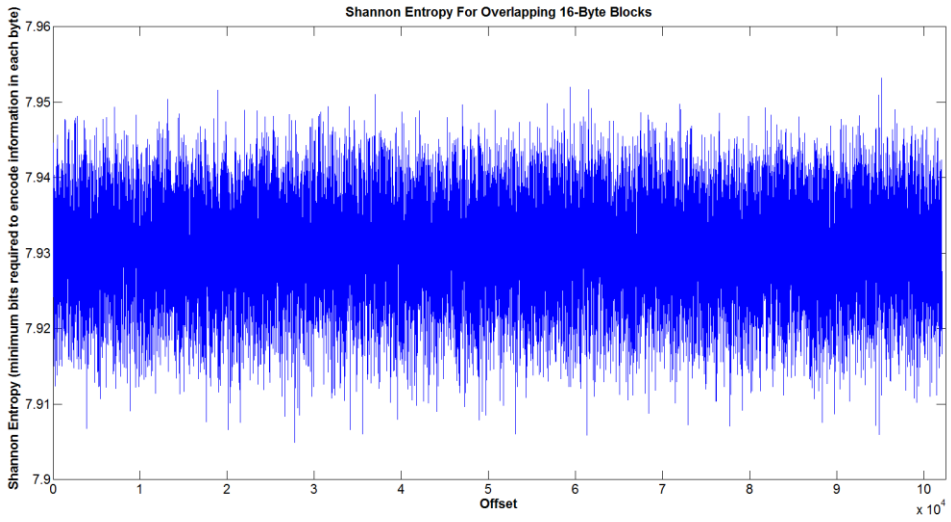


Why this is occurring in the data can only be speculated: perhaps an M out of N counter with frequent roll over. More analysis is required.

2.3. SHANNON ENTROPY OF 16-BYTE BLOCKS

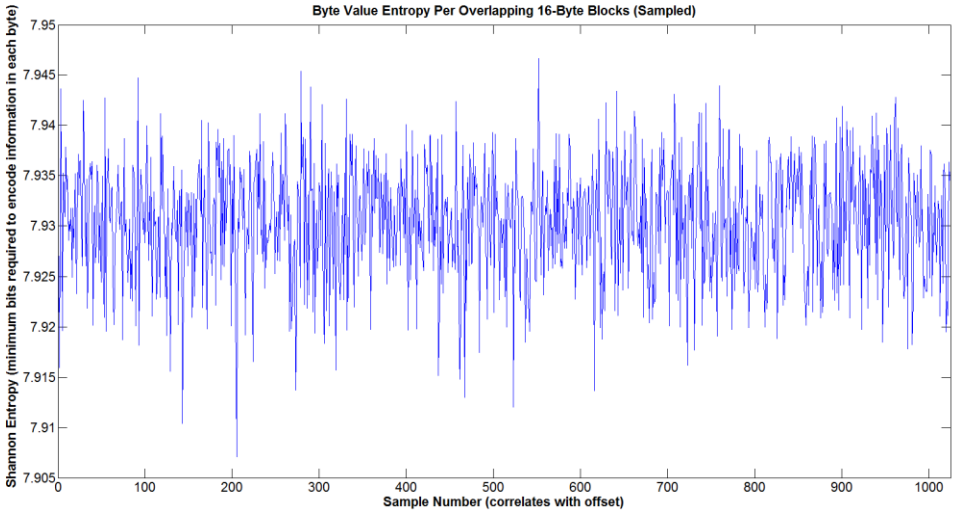
2.3.1. UN-SAMPLED

Byte value Shannon Entropy of contiguous, overlapping 16 byte blocks across all files was computed. This test measures entropy as a function of file offset, and should be able to detect changes in local entropy indicative of chunks of unencrypted data, such as headers. The plotted results look cluttered and noisy; however it's obvious that entropy remains high (average 7.93 with 8 bits per byte representing maximum entropy) throughout.



2.3.2. SAMPLED

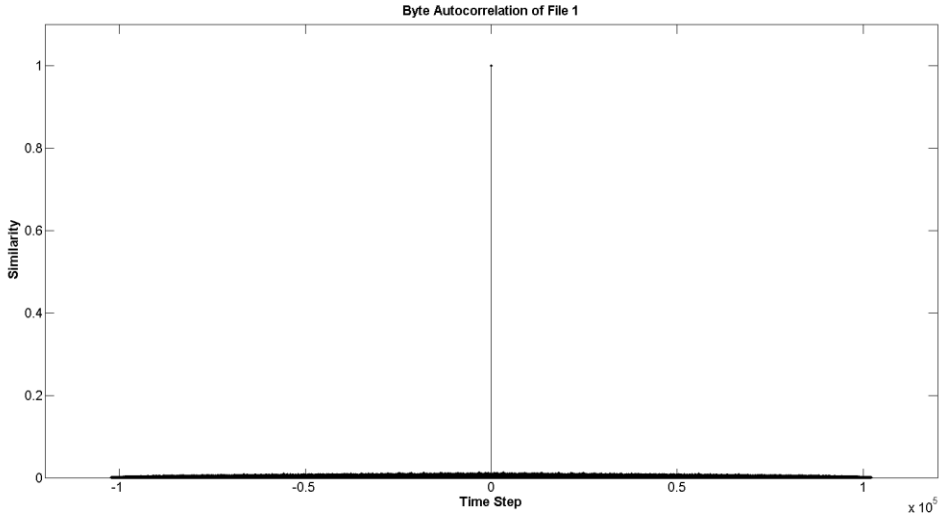
Byte value Shannon Entropy of contiguous, overlapping 16 sample blocks all files when randomly sampled. Each file was randomly sampled, producing 1024 sample elements. The samples from all files were combined, and then the entropy of contiguous, overlapping 16 sample blocks across all files was measured. The results, when plotted, seem to reveal a weak sinusoidal signal which might indicate low entropy at regular regions.



2.4. SIGNAL ANALYSIS

2.4.1. AUTOCORRELATION

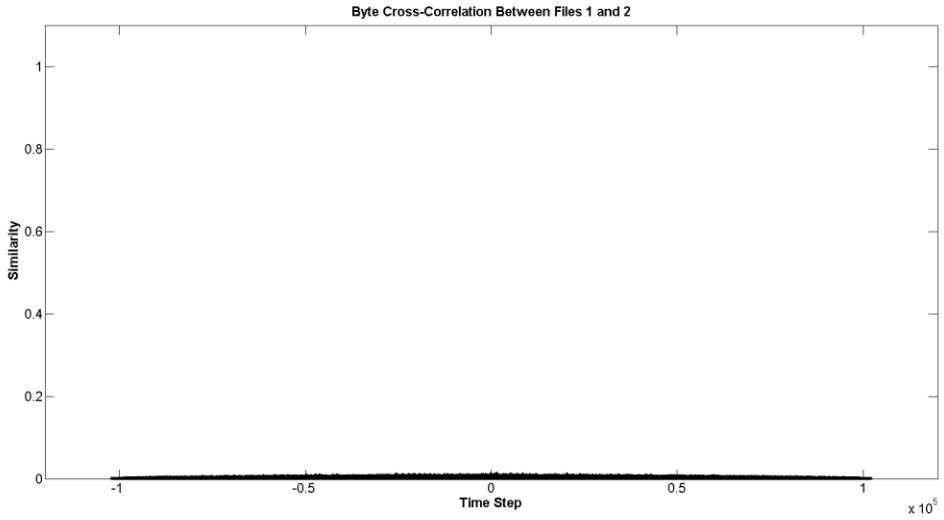
This test measures the similarity of a signal with itself as a function of time. If the data obscures an underlying repetitive signal, this test should detect it if significant. However, the following plot demonstrates that aside from the spike at 0 (the signal is always identical to itself at time step 0); no other similarities were found at different time steps.



The same test was conducted for all files in the data set with similar results.

2.4.2. CROSSCORRELATION

This test measures similarity of a signal with another. If patterns of data are recurring in each file or directory at different offsets, this test would show peaks above noise at the value of offset pertinent to the relative shift. No significant peaks with any of the files cross-correlated with any others were evident.



3. CONCLUSION

In vitro analysis of the encoded data set showed no revealing patterns or indications of information leakage. File number and size patterns provided plausible explanations regarding the source data and might also indicate the use of standard block encryption standards. Signal and statistical analysis confirmed a perfectly random pattern to the data and high degree of entropy that one would expect from randomization produced by encryption algorithms.

It is A^2I's experience in data and network protocol analysis that the origin of encoded data provides a productive means for ascertaining content and coding. Specifically, should the application that generated the presented data become known and openly available as an end user product or of common knowledge, any reverse engineering technique would likely use the product to analyze the structure of output based on known inputs. At least for this study, none of the cipher text data sets presented were paired with their plaintext equivalent input or the application that produced them. Perhaps a productive follow-on step would assume access to the production application. Experiments using known inputs compared to known outputs may shed light on the nature of the data and attempts at reverse engineering and decryption.

4. COMPANY OVERVIEW

Allied Associates International (A^2I) is an internationally recognized expert in the collection, reconstruction and analysis of network protocols with a primary focus on networking communication analysis and exploitation. Founded in 2008, A^2I is a Service Disabled Veteran Owned Small Business (SDVOSB) headquartered in Warrenton, Virginia. We specialize in the development of Cyber Intelligence software tools and solutions for the law enforcement and intelligence communities, and provide a range of mission support solutions for the Department of Defense (DoD). Our employees are computer engineers, scientists, electrical engineers and Subject Matter Experts who have extensive experience working on critical Government programs as former members of the Military, Civil Service and industry.